

Statistické rozhodování a jeho souvislost se ztrátovou funkcí

Statistical decision problem in context with loss function

Zadání diplomové práce

Student:

Bc. Pavel Náplava

Studijní program:

N2647 Informační a komunikační technologie

Studijní obor:

1103T031 Výpočetní matematika

Téma:

Statistické rozhodování a jeho souvislost se ztrátovou funkcí
Statistical decision problem in context with loss function

Zásady pro vypracování:

V práci bude nastíněn statistický rozhodovací problém a jeho řešení v různých praktických situacích v souvislosti s volbou ztrátové funkce. Předpokladem je použití Bayesovského přístupu, jako základního nástroje řešení při různých obměnách ztrátové funkce.

Postup práce:

1. Statistický rozhodovací problém - studium základů.
2. Bayesovský přístup – osvojení metodiky.
3. Metodické řešení základních rozhodovacích úloh při variantních volbách ztrátové funkce.
4. Bezpečnostní ztrátová funkce (Precautionary Loss Function) a její význam v inženýrských úlohách.
5. PC implementace a řešení konkrétních inženýrských úloh.

Seznam doporučené odborné literatury:

- Martz F.H., Waller R.A., Bayesian Reliability Analysis, Wiley1982, ISBN 0-471-86425-0.
- Nostrom J.G., The Use of Precautionary Loss Functions in Risk Analysis, IEEE Transactions on Reliability, vol.45, No. 3, 1996.
- Briš R., Litschmannová M., STATISTIKA II., E-learningový prvek pro podporu výuky odborných a technických předmětů, v rámci projektu CZ.O4.01.3/3.2.15.2/0326, VŠB TU Ostrava, 2007, ISBN 978-80-248-1482-7.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **prof. Ing. Radim Briš, CSc.**

Datum zadání: 01.09.2013

Datum odevzdání: 07.05.2014



doc. RNDr. Jiří Bouchala, Ph.D.
vedoucí katedry





prof. RNDr. Václav Snášel, CSc.
děkan fakulty

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

V Ostravě 7. května 2014


.....

Abstrakt

Bayesovská statistika je moderní disciplína opírající se o na první pohled poměrně základní statistickou větu, a to Bayesovu větu o podmíněné pravděpodobnosti. Bayesovský přístup ke statistice je intuitivní, nicméně je také velice hluboký a rozmanitý. Cílem této práce bylo vytvořit vertikální průřez touto problematikou. Seznámení s rozhodováním za nejistoty, ztrátovými funkcemi, bezpečnostní třídou ztrátových funkcí a numerickou integrací pomocí metod Markov Chain Monte Carlo. A metody simulovaného žíhání pro řešení optimalizačních problémů při rozhodování.

Klíčová slova: Bayes, rozhodovací problém, ztrátová funkce, metody Monte Carlo, Markov Chain Monte Carlo, numerická optimalizace, numerická integrace, simulované žíhání

Abstract

Bayesian statistics is a modern discipline built on, at first look rather rudimentary, the Bayes theorem on conditional probability. Bayesian approach to statistics is intuitive but at the same time very deep and varied. The goal of this thesis was to provide a vertical slice of Bayesian Statistics. That is to introduce decision-making under uncertainty, loss functions, precautionary loss functions and numeric integration via Markov Chain Monte Carlo methods. Present simulated annealing as a method for numerical solution to optimisation problems in decision-making.

Keywords: Bayes, statistical decision theory, loss function, Monte Carlo methods, Markov Chain Monte Carlo, numerical optimisation, numerical integration, simulated annealing

Seznam použitých zkratk a symbolů

$\pi(\Theta)$	– apriorní rozdělení pravděpodobnosti
$\pi(\Theta X)$	– posteriorní rozdělení pravděpodobnosti
$f(X \Theta)$	– funkce věrohodnosti
X	– množina možných hodnot dat
x	– konkrétní datový vzorek/pozorování
Θ	– množina možných hodnot náhodné veličiny
θ	– skutečná hodnota náhodné veličiny (state of the nature)
$\hat{\theta}$	– aproximace θ
MCMC	– Markov Chain Monte Carlo

Obsah

1	Úvod	5
2	Bayesovská statistika	6
2.1	Bayesova věta	6
2.2	Motivace Bayesovského přístupu	7
2.3	Apriorní rozdělení pravděpodobnosti	8
2.4	Aposteriorní rozdělení pravděpodobnosti	9
2.5	Konjugovaný systém rozdělení pravděpodobnosti	9
3	Bayesovský rozhodovací problém	12
3.1	Prvky statistického rozhodovacího problému	12
4	Ztrátové funkce	14
4.1	Vztah užitku ke ztrátě a zavedení uspořádání na množině možných rozhodnutí	14
4.2	Typické ztrátové funkce	15
5	Precautionary loss function - třída ztrátových funkcí	19
5.1	Motivace	19
5.2	Vlastnosti	19
6	Metodika řešení rozhodovacích problémů	22
6.1	MCMC - numerické řešení integrálů	22
7	PC implementace a řešení úloh	27
7.1	Řešená úloha	27
7.2	Implementace algoritmů a numerická řešení	28
8	Závěr	34
9	Reference	35

Seznam tabulek

1	Data ke zkouškám životnosti	27
---	---------------------------------------	----

Seznam obrázků

1	Ilustrace základní zamítací metody	24
2	Ilustrace generalizované zamítací metody	25
3	Ukázka výstupu M-H algoritmu	31
4	Ukázka výstupu M-H algoritmu s úpravou pro simulování žíhání	33

Seznam výpisů zdrojového kódu

1	Ukázka implementace M-H algoritmu v Pythonu	29
2	Ukázka implementace M-H algoritmu s úpravou pro simulování žíhání v Pythonu	31

1 Úvod

Bayesovská statistika je moderní disciplína opírající se o poměrně základní statistickou větu a to Bayesovu o podmíněné pravděpodobnosti. Nicméně její aplikace jsou velice pestré, s velkými přesahy do jiných oborů. Jako například v optickém rozpoznávání znaků, rozpoznávání hlasu, medicínské problémy jako predikce, analýzy rozptylu nebo kategorizace či klinické testování, návrh experimentů, sestavování regresních modelů a tak dále.

Bayesovský přístup ke statistice je poměrně intuitivní, nicméně je také velice hluboký a rozmanitý, jak je ilustrováno v knihách a článcích jako například [7, 8, 11, 13].

Za cíl této práce jsem si kladl vytvořit vertikální průřez touto problematikou, to znamená seznámit čtenáře nejen se základy tohoto přístupu ale i s jeho možnými aplikacemi i v situacích kdy analytické řešení problému není k dispozici. K tomuto účelu jsem zahrnul seznámení s problematikou numerické integrace pomocí metod Markov Chain Monte Carlo. Tyto metody jsou velmi detailně popsány v [9, 12, 14, 15].

Hlavní část tohoto textu je věnována statistickým rozhodovacím problémům a práci J. G. Norstrøma, který v [3, 4] prezentoval třídu asymetrických ztrátových funkcí.

Problematika statistického rozhodování spočívá v minimalizaci nejistot a rizik. Tato část Bayesovské statistiky je detailně popsána ve dvojici knih *Statistical Decision Theory and Bayesian Analysis* (Berger 1988) a *Bayesian Reliability Analysis* (Martz, Waller, 1982) [1, 2], které byly velmi užitečné při tvorbě této práce. Dále pak zadefinujeme bayesovské odhady a odvodíme je pro typické ztrátové funkce. Poté bude prezentována třída bezpečnostních ztrátových funkcí a následovat bude kapitola věnovaná aplikaci prezentované teorie.

Tato práce předpokládá znalost základních vlastností markovských řetězců jako jsou irreducibilita, aperiodicita, stacionární distribuce.

2 Bayesovská statistika

Bayesovská statistika je významnou moderní disciplínou založenou převážně na výsledcích anglického matematika a theologa reverenda Thomase Bayese (1701-1761), především po něm pojmenované Bayesově větě, která nám dává algebraický nástroj pro práci s podmíněnými pravděpodobnostmi. Dílo Thomase Bayese znovuobjevil, rozšířil a zobecnil francouzský matematik a astronom Piére-Simon Laplace (1749-1827).

V Bayesovské statistice je každý zdroj nejistoty chápán jako náhodná veličina a její pravděpodobnost může být interpretována jako stupeň věrohodnosti. Tato metodologie přímo pracuje se subjektivní nebo předem známou informací související s řešeným problémem.

Bayesovský přístup ke statistice je dnes moderní disciplínou, která velmi těží z teoretických poznatků pro numerické řešení integrálů vysokých dimenzí a technologických pokroků na poli výpočetní techniky umožňující jejich numerickou implementaci.

2.1 Bayesova věta

Věta 2.1 *Mějme náhodné jevy A a B s pravděpodobnostmi $P(A)$ a $P(B)$, za předpokladu $P(B) > 0$ platí:*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

$P(A|B)$ je podmíněná pravděpodobnost jevu A za předpokladu že nastal jev B , analogicky $P(B|A)$ je pravděpodobnost jevu B podmíněná výskytem jevu A .

Důkaz. Vyjdeme ze sdružené pravděpodobnosti, která je komutativní:

$$P(A \cap B) = P(B \cap A) \quad (2)$$

Sdruženou pravděpodobnost můžeme psát jako:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \quad (3)$$

$$P(A|B)P(B) = P(B|A)P(A) \quad (4)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (5)$$

■

2.2 Motivace Bayesovského přístupu

Uveďme příklad, na kterém demonstrujeme nutnost zahrnout předem známou nebo subjektivní informaci, do statistického problému. L.J. Savage v [17] uvedl jako příklad tři statistické experimenty:

- Paní, která si přidává do čaje mléko tvrdí, že je schopna pokaždé říct, jestli bylo mléko nalito do šálku před čajem. V deseti pokusech pokaždé uspěla.
- Hudební expert tvrdí, že je schopen pokaždé poznat list not z Haydnova díla od listu not z Mozartova díla. V deseti pokusech pokaždé uspěl.
- Opilý kamarád tvrdí, že je schopen pokaždé předpovědět hod typickou mincí. V deseti pokusech pokaždé uspěl.

Ve všech třech výše uvedených situacích je neznámou veličinou θ pravděpodobnost úspěchu (správné odpovědi) v daném pokusu. Klasický přístup by jako typickou nulovou hypotézu uvažoval, že dotyčný hádá (tj. $H_0 : \theta = 0,5$). Ve všech třech případech je tato H_0 zamítnuta s hladinou spolehlivosti 2^{-10} , takže je velice pravděpodobné, že výše uvedená tvrzení jsou stejně pravdivá.

V druhé situaci je tento závěr snadno přijatelný. V první situaci je možný, nicméně ve třetí situaci cítíme že bychom měli sled deseti úspěšných pokusů pokládat za odlehle pozorování.

Každý asi cítíme, že tyto situace nejsou úplně stejné, nicméně klasický statistický přístup nám neposkytuje možnost naši subjektivní informaci metodicky začlenit do rozhodování o hypotézách týkajících se θ .

Bayesova věta nám dává nástroj k práci se subjektivní pravděpodobností nebo předem dostupnou informací při statistickém rozhodování.

Přeznačme v Bayesově větě o podmíněné pravděpodobnosti $A \Rightarrow H$ jako hypotéza, $B \Rightarrow D$ jako Data.

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)} = \frac{P(D|H)P(H)}{\sum_H (P(D|H)P(H))} \quad (6)$$

- $P(H|D)$: představuje pravděpodobnost (míru důvěryhodnosti) hypotézy při datech (naměřených hodnotách) D - posteriorní rozdělení pravděpodobnosti
- $P(D|H)$: funkce věrohodnosti dat pod danou hypotézou
- $P(H)$: pravděpodobnost (míra důvěryhodnosti) hypotézy před tím než získáme data - apriorní rozdělení pravděpodobnosti
- $P(D)$: normalizační konstanta (pravděpodobnost dat pod jakoukoli hypotézou)

Zatím byla uvedena diskrétní forma bayesovy věty, která nakonec bývá použita k praktickým výpočtům, ukažme i její spojitou formu:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\pi(\theta)} = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta) d\theta} \quad (7)$$

2.3 Apriorní rozdělení pravděpodobnosti

Apriorní rozdělení pravděpodobnosti představuje předem dostupnou nebo subjektivní informaci o rozdělení náhodné veličiny, která je předmětem našeho zájmu.

Jedná se o klíčovou část bayesovského přístupu. Volba apriorního rozdělení není vůbec jednoduchá, jak by se na první pohled mohlo zdát. Pro použití bayesovské statistiky musíme vždy nějaké apriorní rozdělení zvolit, a to i v případech, kdy nemáme žádnou předchozí informaci nebo subjektivní představu o zkoumané náhodné veličině.

Apriorní rozdělení můžeme následovně klasifikovat:

Definice 2.1 Subjektivní apriorní rozdělení

Subjektivní apriorní rozdělení obsahuje předem známou informaci o θ .

Subjektivní apriorní rozdělení lze získat pomocí:

- Histogramu - rozdělením Θ do vhodných intervalů a určením subjektivní pravděpodobnosti, že $\hat{\theta}$ leží v konkrétním intervalu.
- Relativní věrohodnosti - porovnáním intuitivních „věrohodností“ bodů v Θ
- Vybráním vhodné funkční třídy - předpokládáme vhodně vybranou formu funkce se subjektivně volenými konstantami

Problematika volby apriorní distribuce je velmi hluboká, více informací k tomuto tématu lze získat z [1, 2]

Ve snaze použít bayesovský přístup, i v situacích kdy nedisponujeme žádnou nebo pouze minimální informací pro konstrukci apriorního rozdělení, byly vyvinuty takzvané neinformativní nebo objektivní apriorní rozdělení. Slovem neinformativní je myšleno že dané rozdělení vychází pouze z principu neurčitosti.

Definice 2.2 Objektivní apriorní rozdělení

Obsahuje minimální informaci, zpravidla konstruováno na základě principu neurčitosti.

Typicky užívané neinformativní apriorní rozdělení je takzvané Jeffreysovo (1961), které je konstruováno z očekávané Fisherovy míry informace.

Při volbě objektivního nebo neinformativního rozdělení často může dojít k následující situaci:

$$\int \pi(\theta) d\theta = \infty \quad (8)$$

Definice 2.3 Nevlastní apriorní rozdělení

Pro nevlastní rozdělení pravděpodobnosti platí vztah 8.

Tyto rozdělení pak nazýváme nevlastními, protože nesplňují tento základní předpoklad kladený na rozdělení pravděpodobnosti.

2.4 Aposteriorní rozdělení pravděpodobnosti

Aposteriorní rozdělení představuje subjektivní apriorní informaci po konfrontaci s daty. Na základě tohoto rozdělení potom činíme závěry nebo sestavujeme odhady.

Definice 2.4 *Aposteriorní rozdělení pravděpodobnosti*

Aposteriorní rozdělení budeme značit $\pi(\theta|x)$. θ a x mají sdruženou hustotu:

$$h(x, \theta) = f(x|\theta)\pi(\theta) \quad (9)$$

Máme marginální hustotu X :

$$f(x) = \int_{\theta \in \Theta} f(x|\theta)\pi(\theta)d\theta \quad (10)$$

Pro normalizační konstantu $f(x) \neq 0$ dostaneme aposteriorní hustotu následovně:

$$\pi(\theta|x) = \frac{h(x, \theta)}{f(x)} = \frac{f(x|\theta)\pi(\theta)}{\int_{\theta \in \Theta} f(x|\theta)\pi(\theta)d\theta} \quad (11)$$

2.5 Konjugovaný systém rozdělení pravděpodobnosti

Definice 2.5 *Konjugovaný systém rozdělení pravděpodobnosti*

Nechť \mathcal{F} je třída funkcí hustoty $f(x|\theta)$ indexovaných podle θ . Třída \mathcal{P} rozdělení funkcí $\pi(\theta)$ (apriorní) je konjugovaný systém pokud má stejný tvar jako rozdělení z \mathcal{F} pro $\forall \pi(x|\theta) \in \mathcal{F}$ a $\forall \pi(\theta) \in \mathcal{P}$.

$$\mathcal{P} \text{ konjugovaná k } \mathcal{F} \iff (\forall f(x|\theta) \in \mathcal{F} \wedge \forall \pi(\theta) \in \mathcal{P}) : \pi(\theta, x) \in \mathcal{P}$$

Jako příklady konjugovaného systému můžeme uvést množinu všech možných rozdělení pravděpodobnosti, rozdělení exponenciální třídy, dvojice binomické rozdělení a beta rozdělení, dvojice poissonovo rozdělení a gamma rozdělení. Při práci s konjugovaným systémem dochází pouze k „aktualizování“ konstant pro dané rozdělení a také obecně nemusíme explicitně počítat s normalizační konstantou $f(x)$.

Příklad 2.1

Práce s konjugovaným systémem.

Byla provedena zkouška životnosti jisté komponenty letadla. Ze 13 vzorků byl získán následující datový soubor (při 10. poruše byla zkouška ukončena).

0,22 0,50 0,88 1,00 1,32 1,33 1,54 1,76 2,50 3,00

1. Za předpokladu, že se životnost řídí exponenciálním modelem, nalezněte funkci věrohodnosti pro parametr λ .
2. Předpokládejte apriorní rozdělení pro parametr λ ve tvaru $\pi(\lambda) = \alpha \lambda e^{-4\lambda}$

3. Určete odhad parametru λ aposteriorním modelem i aposteriorní očekávanou hodnotou.
4. Spočítejte aposteriorní rozptyl.

Funkce věrohodnosti pro parametr λ

Máme zadáno 10 hodnot měření $x_{(1)}, \dots, x_{(10)}$ a 3 cenzorované vzorky u nichž pouze máme informaci, že během zkoušky životnosti u nich nedošlo k poruše. Jedná se tedy o cenzurování počtem poruch do ukončení experimentu.

Použijeme exponenciální model s parametrem $\lambda \in (0, \infty)$.

Nejdříve sestavíme funkci věrohodnosti pro cenzurování II. druhu:

$$L(x) = \frac{n!}{(n-r)!} \prod_{i=1}^r f(x_{(i)}) \cdot R^{n-r}(x_{(r)})$$

Ad (1) Naši funkci věrohodnosti cenzurování II. druhu dostaneme:

$$L(\lambda; x) = \frac{n!}{(n-r)!} \cdot \prod_{i=1}^r f(x_{(i)}) \cdot R^{n-r}(x_{(r)})$$

$$L(\lambda; x) = \frac{13!}{(13-10)!} \cdot \prod_{i=1}^{10} \lambda e^{-\lambda x_{(i)}} \cdot R^{13-10}(x_{(10)})$$

$$\text{kde } R(x) = 1 - F(x)$$

$$L(\lambda; x) = \frac{13!}{3!} \cdot \prod_{i=1}^{10} \lambda e^{-\lambda x_{(i)}} \cdot (1 - (1 - e^{-\lambda x_{(10)}}))^3$$

$$L(\lambda; x) = \frac{13! \cdot 10}{3!} \cdot \lambda^{10} e^{-\lambda 10\bar{x}} \cdot e^{-\lambda 3x_{(10)}} \propto \lambda^{10} e^{-\lambda 23,05}$$

Aposteriorní rozdělení λ

Nyní sestavíme funkci aposteriorního rozdělení parametru $\lambda \in (0, \infty)$.

$$\pi(\lambda; \bar{x}) = \frac{L(\lambda; x) \cdot \pi(\lambda)}{\int_{\lambda} L(\lambda; x) \cdot \pi(\lambda) d\lambda}$$

Z výše uvedeného je vidět, že není třeba se zabývat konstantami nebo proměnnými bez závislosti na λ , neboť se po vytknutí z integrálu ve jmenovateli zkrátí.

Integrál ve jmenovateli hraje roli normalizační konstanty (pro splnění $\int f(x) dx = 1$)

Ad (2) Apriorní rozdělení λ :

$$\pi(\lambda) = \alpha \lambda e^{-4 \cdot \lambda}$$

Takže:

$$\text{aposteriorní rozdělení} \propto \lambda e^{-4\lambda} \cdot \lambda^{10} e^{-\lambda 23,05}$$

$$\pi(\lambda|x) \propto \lambda^{11} e^{-\lambda 27,05}$$

A normalizační konstantu:

$$\int_{\lambda} \lambda^{11} e^{-\lambda 27,05} d\lambda = 2,60105 \cdot 10^{-10}$$

■

Poznámka 2.1 Tímto příkladem jsme chtěli ilustrovat, jak nám může konjugovaný systém usnadnit práci, v příkladu došlo pouze k aktualizaci parametrů rozdělení.

3 Bayesovský rozhodovací problém

Statistické rozhodování je obor statistiky zabývající se navrhováním technik pro rozhodování v situacích, kdy je nutno zohlednit stochastické komponenty problematiky, jinými slovy rozhodování za přítomnosti nejistoty.

Aplikace výsledků na poli statistického rozhodování jsou rozsáhlé, od testování hypotéz, bodových odhadů parametrů, hledání konfidenčních intervalů po aplikace například v business intelligence nebo i optickém rozpoznávání znaků.

Poprvé tento obor vyvinul Wald (1950). Bayesovskou teorii rozhodování pak Box a Tiao (1973), Morgan (1968).

3.1 Prvky statistického rozhodovacího problému

1. množina Θ , která obsahuje všechny možné skutečné hodnoty θ
2. apriorní rozdělení pravděpodobnosti náhodné veličiny $\Theta : \pi(\theta)$
3. pozorovatelná náhodná veličina X , taková, že pokud je θ skutečný stav systému, pak tato náhodná veličina má rozdělení pravděpodobnosti $f(\theta|x)$
4. množina akcí \mathcal{A} obsahující všechny akce a které mají pro daný problém smysl
5. množina rozhodovacích funkcí \mathcal{D} , které jsou definované na X a zobrazují ji do \mathcal{A}
6. ztrátová funkce $L(\theta, a)$ představující ztrátu plynoucí z vykonání akce a při „skutečném“ stavu θ

Pro konkrétní rozhodovací funkci $\delta \in \mathcal{D}$ máme ztrátovou funkci:

$$L(\theta, \delta(x)) \quad (12)$$

Je tedy vidět, protože rozhodovací funkce má jako parametr datový vzorek, že ztráta závisí na získaných datech a je náhodná veličina. Takže má smysl uvažovat její očekávanou hodnotu, kterou pojmenujeme riziko:

$$R(\theta, \delta) = E(L(\theta, \delta(X))|\theta) = \int_{x \in X} L(\theta, \delta(x)) f(x|\theta) dx \quad (13)$$

Dobré rozhodnutí bude ideálně minimem rizika pro všechna $\theta \in \Theta$. Fakt, že hodnotu $\hat{\theta}$ neznáme, omezuje použitelnost rizika jako kritéria pro výběr optimální rozhodovací funkce, nicméně nám na množině \mathcal{D} zavádí uspořádání pro vybraná θ . Takže na základě porovnání rizik můžeme upřednostňovat výběr rozhodovací funkce s menším rizikem.

Definice 3.1 Bayesovské riziko je očekávaná hodnota rizika při apriorním rozdělení $\pi(\theta)$

$$r(\pi(\theta), \delta) = E(R(\Theta, \delta)) = \int_{\theta \in \Theta} \int_{x \in X} L(\theta, \delta(x)) f(x|\theta) \pi(\theta) dx d\theta \quad (14)$$

Bayesovské riziko má větší výpovědní hodnotu o kvalitě dané rozhodovací funkce, protože na \mathcal{D} zavádí lineární uspořádání.

Definice 3.2 Bayesovská rozhodovací funkce je taková rozhodovací funkce jež, pokud existuje, představuje minimalizaci bayesovského rizika pro dané rozdělení $\pi(\theta)$.

$$r(\pi(\theta)) = r(\pi(\theta), \delta) = \min_{\delta \in \mathcal{D}} r(\pi(\theta), \delta)$$

A její asociované riziko nazýváme minimální bayesovské riziko.

Poznámka 3.1 Bayesovská rozhodovací funkce vždy vybírá akci s minimálním aposteriorním rizikem, této akci říkáme bayesovský odhad nebo bayesovský estimátor.

Konstrukce bayesovské rozhodovací funkce.

Za předpokladu, že můžeme zaměnit pořadí integrálů, dostaneme:

$$r(\pi(\theta), \delta) = \int_{x \in X} f(x) \left(\int_{\theta \in \Theta} L(\theta, \delta(x)) \pi(\theta|x) d\theta \right) dx \quad (15)$$

Minimalizace rizika tedy docílíme minimalizací vnitřního integrálu pomocí vhodné volby $\delta(x)$. Vnitřní integrál představuje podmíněnou očekávanou hodnotu $E(L(\Theta, \delta(x))|x)$ ztráty vzhledem k aposteriornímu riziku. Takže bayesovskou rozhodovací funkci lze vybrat bez nutnosti vyčíslení minimálního bayesovského rizika. Označme aposteriorní riziko $\phi(\delta, x)$, pak:

$$\phi(x) = \min_{\delta \in \mathcal{D}} (\phi(\delta, x)) = \min_{\delta \in \mathcal{D}} E(L(\Theta, \delta(x))|x) \quad (16)$$

minimální bayesovské riziko bude dáno:

$$r(\pi(\theta)) = \int_{x \in X} \phi(x) f(x) dx = E(\phi(X)) \quad (17)$$

což je očekávaná hodnota $\phi(X)$ při sdruženém marginálním rozdělení X .

4 Ztrátové funkce

Při kvantifikování míry užitku (utility) důsledků možných akcí vyvstávají dva problémy:

- Nemusí být zcela jednoznačné jaké míry nebo hodnoty ztráty volit (například kvantifikace dobré vůle zákazníka)
- I pokud je k dispozici míra, podle které lze důsledky rozhodnutí hodnotit, nemusí korespondovat se skutečnou hodnotou.

Dále uvedeme metodu pro určení skutečné hodnoty, která povede k zavedení ztrátové funkce.

4.1 Vztah užitku ke ztrátě a zavedení uspořádání na množině možných rozhodnutí

Užitek a funkce užitku (utility function) jsou ekonomické termíny související se ztrátovou funkcí. Uvedeme jeho využití a jeho vztah ke ztrátové funkci. Pomocí užitku můžeme na \mathcal{D} zavést preferenční uspořádání.

$$a_1 \preceq a_2 \iff U(\theta, a_1) \leq U(\theta, a_2)$$

Slovy: akci a_2 preferujeme před akcí a_1 .

Vztah mezi funkcí užitku a ztrátovou funkcí:

$$U(\theta, a) = -L(\theta, a) \quad (18)$$

Připomeneme rozhodovací problém:

$$\phi(x) = \min_{\delta \in \mathcal{D}} (\phi(\delta, x)) = \min_{\delta \in \mathcal{D}} E(L(\Theta, \delta(x)) | x)$$

$\pi(\theta | x)$ je aposteriorní rozdělení θ pro dané x . Při dané ztrátové funkci a daném aposteriorním rozdělení existuje přirozené uspořádání $a_1 \preceq a_2$, slovy: akce a_2 je alespoň tak dobrá jako a_1 .

$$a_1 \preceq a_2 \iff \int L(\theta, a_1) \pi(\theta) d\theta \geq \int L(\theta, a_2) \pi(\theta) d\theta$$

Pokud disponujeme daty a máme preferenční vztah mezi akcemi a_1 a a_2 můžeme ztrátovou funkci konstruovat pomocí funkce věrohodnosti dat.

$$a_1 \preceq a_2 \iff \int L(\theta, a_1) \pi(\theta | x) d\theta \geq \int L(\theta, a_2) \pi(\theta | x) d\theta$$

Uspořádání můžeme také samozřejmě postavit na bayesovském riziku:

$$a_1 \preceq a_2 \iff r(\pi, a_1) \geq r(\pi, a_2)$$

Na základě výše zmíněných kritérií můžeme zkonstruovat preferenční uspořádání, a tudíž i k němu přidruženou ztrátovou funkci, na libovolné množině.

4.2 Typické ztrátové funkce

Pro přehlednost odvození bayesovských odhadů se nyní omezíme na rozhodování v kontextu odhadu parametru, to pro nás znamená $X \subset \mathcal{R}^n$, $n \in \mathcal{N}$, $n \geq 1$, $\mathcal{D} \equiv X$, $\delta(x) = \theta$.

Definice 4.1 Binární (0-1) ztrátová funkce

$$L(\hat{\theta}, \theta) = 1 - \delta_{\hat{\theta}\theta} = \begin{cases} 0 & \text{pro } \hat{\theta} = \theta \\ 1 & \text{pro } \hat{\theta} \neq \theta \end{cases}$$

Kde $\delta_{..}$ je Kroneckerova Delta funkce:

$$\delta_{ij} = \begin{cases} 1 & \text{pro } i = j \\ 0 & \text{pro } i \neq j \end{cases}$$

Tato ztrátová funkce může být použita při práci v diskrétním prostoru akcí, například v klasifikačním problému za předpokladu, že každá špatná klasifikace způsobuje stejnou ztrátu.

Očekávaná ztráta při této funkci je:

$$\begin{aligned} R(\hat{\theta}) &= \sum_{\mathcal{D}} L(\hat{\theta}, \theta) P(\theta = \hat{\theta}|x) \\ &= \sum_{\mathcal{D}} (1 - \delta_{\hat{\theta}\theta}) P(\theta = \hat{\theta}|x) \\ &= \sum_{\mathcal{D}} P(\theta = \hat{\theta}|x) - \sum_{\mathcal{D}} \delta_{\hat{\theta}\theta} P(\theta = \hat{\theta}|x) \\ &= 1 - P(\theta = \hat{\theta}|x) \end{aligned} \tag{19}$$

Toto riziko dosahuje svého minima pokud zvolíme $\hat{\theta}$ jako maximum aposteriorního rozdělení, tomuto odhadu říkáme MAP (maximum a posteriori).

Definice 4.2 Kvadratická ztrátová funkce

$$L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$$

Na rozdíl od binární ztrátové funkce je v tomto případě zohledněna i míra rozdílu mezi odhadem a skutečnou hodnotou parametru. Tato ztrátová funkce je vhodná, pokud Θ je spojitá s dobře definovanou metrikou.

$$\begin{aligned} R(\hat{\theta}) &= \int L(\hat{\theta}, \theta) f(\theta|x) d\theta \\ &= \int (\hat{\theta} - \theta)^2 f(\theta|x) d\theta \\ &= \hat{\theta}^2 \int f(\theta|x) d\theta - 2\hat{\theta} \int \theta f(\theta|x) d\theta + \int \theta^2 f(\theta|x) d\theta \\ &= \hat{\theta}^2 - 2\hat{\theta} E(\theta) + E(\theta^2) \\ &= \hat{\theta}^2 - 2\hat{\theta} E(\theta) + E(\theta)^2 + E((\theta - E(\theta))^2) \\ &= (\hat{\theta} - E(\theta))^2 + E((\theta - E(\theta))^2) \end{aligned} \tag{20}$$

Všechny očekávané hodnoty jsou brány podle $f(\theta|x)$. Druhá část součtu na posledním řádku rovnice 20 nezávisí na $\hat{\theta}$. K minimalizaci rizika tedy dochází pokud je $\hat{\theta}$ voleno jako očekávaná hodnota θ podle aposteriorního rozdělení pravděpodobnosti. Minimální ztráta je pak dána kovariancí aposteriorního rozdělení.

Definice 4.3 *Absolutní ztrátová funkce*

$$L(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$$

Riziko při této ztrátové funkci:

$$\begin{aligned} R(\hat{\theta}) &= \int L(\hat{\theta}, \theta) f(\theta|x) d\theta \\ &= \int |\hat{\theta} - \theta| f(\theta|x) d\theta \\ &= \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) f(\theta|x) d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) f(\theta|x) d\theta \end{aligned} \quad (21)$$

Integrací metodou per partes lze ukázat že:

$$\begin{aligned} \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) f(\theta|x) d\theta &= \int_{-\infty}^{\hat{\theta}} F(\theta < y|x) dy \\ \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) f(\theta|x) d\theta &= \int_{\hat{\theta}}^{\infty} F(\theta > y|x) dy \end{aligned} \quad (22)$$

Kde $F(\theta < y|x)$ a $F(\theta > y|x)$ jsou kumulativní distribuční funkce.

$$R(\hat{\theta}) = \int_{-\infty}^{\hat{\theta}} F(\theta < y|x) dy + \int_{\hat{\theta}}^{\infty} F(\theta > y|x) dy \quad (23)$$

Derivací podle $\hat{\theta}$ dostaneme minimum v $F(\theta < y|x) = F(\theta > y|x)$, což znamená, že bayesovským odhadem je v tomto případě medián aposteriorního rozdělení pravděpodobnosti.

Příklad 4.1

Pokračování příkladu 2.1.

Připomeňme zadání:

Byla provedena zkouška životnosti jisté komponenty letadla. Ze 13 vzorků byl získán následující datový soubor (při 10. poruše byla zkouška ukončena).

0,22 0,50 0,88 1,00 1,32 1,33 1,54 1,76 2,50 3,00

1. Za předpokladu, že se životnost řídí exponenciálním modelem, nalezněte funkci věrohodnosti pro parametr λ .
2. Předpokládejte apriorní rozdělení pro parametr λ ve tvaru $\pi(\lambda) = \alpha \lambda e^{-4\lambda}$
3. Určete odhad parametru λ aposteriorním modem i aposteriorní očekávanou hodnotou.
4. Spočtěte aposteriorní rozptyl.

Odhad λ pomocí apos. modu a apos. očekávané hodnoty

Ad (3) Máme:

$$\pi(\lambda|x) \propto \lambda^{11} e^{-\lambda 27,05}$$

$$\int_{\lambda} \lambda^{11} e^{-\lambda 27,05} d\lambda = 2,60105 \cdot 10^{-10}$$

Medián a modus jsou tedy:

$$E(\Lambda) = \int_{\lambda} \frac{\lambda \cdot \lambda^{11} e^{-\lambda 27,05}}{\int_{\lambda} \lambda^{11} e^{-\lambda 27,05} d\lambda} d\lambda = 0,443623$$

Modus dostaneme jako řešení rovnice (není třeba normalizační konstanty):

$$\frac{d(\lambda^{11} e^{-\lambda 27,05})}{d\lambda} = 0 \quad \text{pro } \lambda \in (0, \infty)$$

$$\lambda = 0,406654$$

Aposterioerní rozptyl Ad (4) Rozptyl Λ dostaneme z:

$$D(\Lambda) = E(\Lambda)^2 - (E(\Lambda))^2$$

$$E(\Lambda) = \int_{\lambda} \frac{\lambda \cdot \lambda^{11} e^{-\lambda 27,05}}{\int_{\lambda} \lambda^{11} e^{-\lambda 27,05} d\lambda} d\lambda$$

$$E(\Lambda)^2 = \int_{\lambda} \frac{\lambda^2 \cdot \lambda^{11} e^{-\lambda 27,05}}{\int_{\lambda} \lambda^{11} e^{-\lambda 27,05} d\lambda} d\lambda$$

$$D(\Lambda) = 0,0164002$$

■

Poznámka 4.1 Poznamenejme nakonec, že oba odhady λ z příkladu výše:
 Odhad pomocí aposteriorní očekávané hodnoty:

$$E(\Lambda) = \int_{\lambda} \frac{\lambda \cdot \lambda^{11} e^{-\lambda 27,05}}{\int_{\lambda} \lambda^{11} e^{-\lambda 27,05} d\lambda} d\lambda = 0,443623$$

Odhad pomocí aposteriorního modu:

$$\frac{d(\lambda^{11} e^{-\lambda 27,05})}{d\lambda} = 0 \quad \text{pro } \lambda \in (0, \infty)$$

$$\lambda = 0,406654$$

Jsou podle volby metriky ztráty optimální, odhad očekávanou hodnotou v případě kvadratické ztrátové funkce a odhad modem v případě volby ztrátové funkce jako absolutní hodnoty rozdílu.

5 Precautionary loss function - třída ztrátových funkcí

5.1 Motivace

Doposud uvedené ztrátové funkce byly vždy symetrické ve smyslu, že nezáleželo, na které straně náš odhad/rozhodnutí leží, v porovnání se skutečnou hodnotou parametru.

Dokážeme si jistě ale představit, že v reálných aplikacích budeme chtít rozhodovací proces navrhnout, tak abychom například regulovali počet falešně pozitivních vyhodnocení při kategorizačním problému.

5.2 Vlastnosti

Definice 5.1 Ztrátová funkce je pro nějaké $\epsilon > 0$

- *downside damaging* pokud $L(\hat{\theta} - \epsilon, \hat{\theta}) \geq L(\hat{\theta} + \epsilon, \hat{\theta})$ (nechceme podcenit)
- *upside damaging* $L(\hat{\theta} - \epsilon, \hat{\theta}) \leq L(\hat{\theta} + \epsilon, \hat{\theta})$ (nechceme přecenit)
- *symetrická* pokud je zároveň *downside damaging* a *upside damaging*

Definice 5.2 $L(\theta, a)$ je bezpečnostní (precautionary) ztrátová funkce právě tehdy když:

1. $L(\theta, a)$ je *downside damaging*
2. pro každé θ , $L(\theta, a) \rightarrow \infty$ pro $a \rightarrow 0$.

Definice 5.3 Třída bezpečnostních (precautionary) ztrátových funkcí:

$$L(\theta, a) = (\log(a/\theta))^2$$

Věta 5.1 Pro $L(\theta, a)$ z definice 5.3. Je bayesovský odhad dán:

$$\hat{\theta}_p^2 = E_{\pi(\theta|x)}(\theta^2) = \int_{\theta \in \Theta} \theta^2 \pi(\theta|x) d\theta \quad (24)$$

$$\begin{aligned} \hat{\theta}_p &= f \cdot E_{\pi(\theta|x)}(\theta) \\ f^2 &\equiv 1 + \frac{\text{Var}_{\pi(\theta|x)}(\theta)}{E_{\pi(\theta|x)}^2(\theta)} > 1 \end{aligned} \quad (25)$$

Výše uvedená věta je dokázána v [3]

Ztrátovou funkci v 5.3 můžeme zobecnit na tvar:

$$L(\theta, a) = \frac{(\theta - a)^2}{a^k} \cdot w(\theta), 0 < k \leq 2, w(\theta) > 0; \quad (26)$$

kde $w(\theta)$ je libovolná váhová funkce. $k \leq 2$ zaručuje, že ztráta roste s rozdílem $a - \theta$, k je bezpečnostní (precautionary) index, který reguluje míru asymetričnosti ztrátové funkce. V limitě $k \rightarrow 0$ se jedná o klasickou kvadratickou ztrátu, pro $k = 2$ se blíží k $w(\theta)$ pro $a \rightarrow \infty$.

Možný způsob volby konstanty k :

- vybereme dvě hodnoty $a, \theta - \epsilon$, a $\theta + c \cdot \epsilon$, $c > 1$, jež nesou identickou ztrátu
- řešíme $L(\theta, \theta - \epsilon) = L(\theta, \theta + c \cdot \epsilon)$ pro k

Zvolme $\epsilon = \frac{\theta}{c}$:

$$L(\theta, \theta - \frac{\theta}{c}) = L(\theta, 2\theta) \quad (27)$$

Řešením pro k získáme:

$$k = \frac{2 \log(c)}{\log\left(\frac{2c}{(c-1)}\right)} \quad (28)$$

Dva speciální případy při volby konstanty c :

- $c = 2$ dává $k = 1$
- $c = 3$ dává $k = 2$

Věta 5.2 Při ztrátové funkci uvedené ve (26) je bayesovský estimátor dán:

$$\hat{\theta}_p = \frac{1}{\Psi_1(x)} ((1-k) + \sqrt{(1-k)^2 + \Psi_1(x)\Psi_2(x)})$$

$$\Psi_1(x) \equiv (2-k) \frac{E_{\pi(\theta|x)}(w(\theta))}{E_{\pi(\theta|x)}(\theta w(\theta))}$$

$$\Psi_2(x) \equiv k \frac{E_{\pi(\theta|x)}(\theta^2 w(\theta))}{E_{\pi(\theta|x)}(\theta w(\theta))}$$

Větu 5.2 lze dokázat derivováním aposteriorní očekávané ztráty a položením rovno 0. Posléze řešením pro a identifikujeme kritické body.

Větu 5.1 dostaneme z 5.2 při $k = 1$ a $w(\theta) = 1$.

Uvedeme příklad z [4]

Příklad 5.1

Nostrøm prezentuje použití bezpečnostní ztrátové funkce následovně:

Pro data o spolehlivosti z ropných plošin odpovídající poissonovu modelu, tedy r selhání v čase t . Konjugovaná apriorní hustota je pro tento případ dána jako funkce gamma:

$$\pi(\lambda|\alpha, \beta) = \beta \cdot \Gamma(\beta\lambda; \alpha) \quad (29)$$

A podle bayesovy věty máme tedy aposteriorní rozdělení dáno jako:

$$\pi(\lambda|r; t, \alpha, \beta) = (t + \beta) \cdot \Gamma((t + \beta)\lambda; r + \alpha) \quad (30)$$

při pozorování r selhání v čase t . Pro konstrukci odhadu $\lambda \equiv \theta$ potřebujeme aposteriorní očekávanou hodnotu počtu selhání a aposteriorní rozptyl.

$$\begin{aligned}\hat{\lambda}_k &= \frac{(s + \alpha)}{(t + \beta)} \\ \text{Var} &= \frac{(s + \alpha)}{(t + \beta)^2}\end{aligned}\tag{31}$$

Kvadratická ztrátová funkce nám dává $\hat{\lambda}$ jako bayesovský estimátor, který srovnáme s odhadem získaným pomocí bezpečnostní ztrátové funkce s koeficientem $k = 1$.

$$\begin{aligned}\hat{\lambda}_b &= \frac{\sqrt{(s + \alpha) \cdot (s + \alpha + 1)}}{(t + \beta)} \\ &= \hat{\lambda}_k \cdot f \\ f &\equiv \sqrt{1 + \frac{1}{s + \alpha}}\end{aligned}\tag{32}$$

Na základě znalosti koeficientů aposteriorního Γ rozdělení

$$\begin{aligned}r + \alpha &= 1,31 \\ t + \beta &= 25300\end{aligned}\tag{33}$$

určil:

$$\begin{aligned}\hat{\lambda}_q &= 51,7 \cdot 10^{-6} / \text{hodina} \\ \text{Var} &= 20,5 \cdot 10^{-10} / \text{hodina}^2\end{aligned}\tag{34}$$

Bayesovský estimátor při ztrátové funkci pak je:

$$\begin{aligned}\hat{\lambda}_b &= \hat{\lambda}_k \cdot f = 1,33 \cdot 51,7 \cdot 10^{-6} / \text{hodina} \\ &= 68,7 \cdot 10^{-6} / \text{hodina} = 0,60 / \text{rok}\end{aligned}\tag{35}$$

Je vidět že odhad získaný při použití bezpečnostní ztrátové funkce je v tomto případě zhruba o 30 % vyšší, oproti odhadu při použití kvadratické ztrátové funkce $\hat{\lambda}_q$.

■

6 Metodika řešení rozhodovacích problémů

V předchozí kapitole jsme odvodili bayesovské odhady pro typické a bezpečnostní ztrátové funkce, mohou však nastat i situace, ve kterých není jednoduché tyto odhady získat analyticky, a proto má smysl v takovém případě přistoupit k numerickému řešení. Tato problematika je detailněji diskutována v knihách a článcích [9, 12, 14, 15], ze kterých jsem také čerpal.

Integrály, se kterými se setkáváme v bayesovském přístupu, často bývají velice složité pro analytické řešení.

$$\int_{x_1, \dots, x_n} g(\theta) f(\theta|x) dx_1 \dots dx_n \quad (36)$$

Často se tedy musíme spokojit s jejich numerickým řešením, které je navíc komplikováno vysokou dimenzí zmíněných integrálů.

6.1 MCMC - numerické řešení integrálů

Řešení těchto integrálů se nabízí ve formě simulačních metod, takzvaných Monte Carlo metod.

Cílem těchto metod je výpočet integrálu pomocí aproximace součtem na základě silného zákona velkých čísel.

$$I(g) = E_f(g(X)) = \int g(x) f(x) dx \quad (37)$$

Za předpokladu, že umíme získat nezávislé vzorky $x^{(1)}, \dots, x^{(n)}$ z rozdělení $f(x)$, můžeme pak aproximovat:

$$\int g(x) f(x) dx \approx \frac{1}{N} \sum_{i=1}^N g(x^{(i)}) \quad (38)$$

6.1.1 Generování vzorků z rozdělení náhodné veličiny

Základem všech metod pro generování vzorků z libovolného rozdělení je kvalitní generátor pseudonáhodných čísel z rovnoměrného rozdělení, jako například Mersene Twister algoritmus (Makoto Matsumoto, Takuji Nishimura 1998).

Věta 6.1 *Integrální transformace pravděpodobnosti*

Nechť je F libovolná distribuční funkce a F^- je její pseudoinverze podle vztahu:

$$F^-(u) = \inf_x (F(x) \geq u) \quad (39)$$

Pak za předpokladu $U \sim \mathcal{U}_{(0,1)}$ pak $F^-(U)$ má distribuční funkci F .

Obecně je integrální transformace málo efektivní. Pro generování vzorků z konkrétních rozdělení se využívají transformace náhodných veličin.

6.1.2 Zamítací metoda

Myšlenkou zamítací metody je:

$$f(x) = \int_0^{f(x)} du \quad (40)$$

Věta 6.2 Základní věta simulace

Simulace $X \sim f(x)$ je ekvivalentní simulací:

$$(X, U) \sim \mathcal{U}((x, u) : 0 < u < f(x)).$$

Při generování (X, U) pracujeme na oblasti $\langle a, b \rangle$ a předpokládáme:

$$\begin{aligned} \int_a^b f(x) dx &= 1 \\ f(x) &\leq m \text{ na } \langle a, b \rangle \end{aligned} \quad (41)$$

Algoritmus zamítací metody:

1. generujeme kandidáta $(Y, U) \sim \mathcal{U}_{\langle a, b \rangle \times \langle 0, m \rangle}$ tj. $Y \sim \mathcal{U}_{\langle a, b \rangle}$, $U|Y \sim \mathcal{U}_{\langle 0, m \rangle}$
2. kandidát na $X \sim f(x)$ přijmeme pokud: $u < f(x)$, jinak kandidáta zahazujeme

Pro vzorky generované tímto postupem platí:

$$\begin{aligned} P(X \leq x) &= P(Y \leq x | U \leq f(Y)) \\ &= \frac{\int_a^x \int_0^{f(y)} du dy}{\int_a^b \int_0^{f(y)} du dy} \\ &= \int_a^x f(y) dy \end{aligned} \quad (42)$$

To znamená, že náhodná veličina X má hustotu pravděpodobnosti $f(x)$.

Zamítací metodu lze zobecnit pro libovolnou oblast. Podmínkou je nutnost umět simulovat z rovnoměrného rozdělení na dané oblasti.

Předpokládáme, že hustotu $f(x)$ lze shora omezit integrovatelnou funkcí $m(x)$.

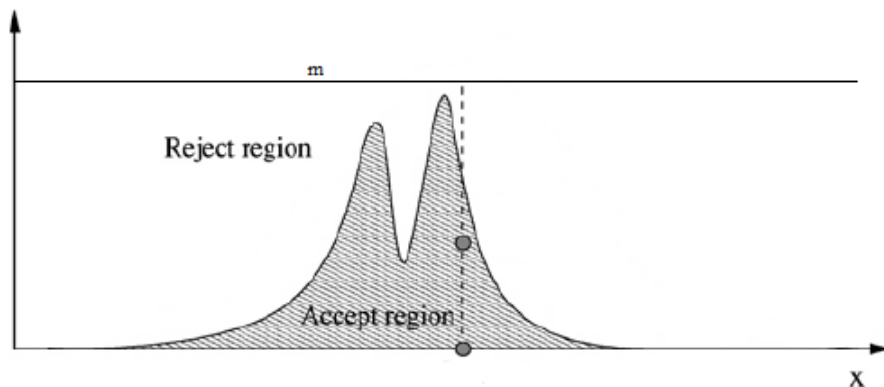
$$M = \int m(x) dx$$

Pak existuje hustota pravděpodobnosti:

$$m(x) = Mg(x)$$

Vektor (Y, U) pak můžeme simulovat po částech:

$$\begin{aligned} Y &\sim g(y) \\ U|Y = y &\sim \mathcal{U}_{\langle 0, Mg(y) \rangle} \end{aligned} \quad (43)$$



Obrázek 1: Ilustrace základní zamítací metody

Opět přijímáme vzorky podle $u < f(y)$.

$$\begin{aligned}
 P(X \in O) &= P(Y \in O | U < f(Y)) \\
 &= \frac{\int_a^x \int_0^{f(y)} \frac{1}{Mg(y)} du g(y) dy}{\int_a^b \int_0^{f(y)} \frac{1}{Mg(y)} du g(y) dy} \\
 &= \int_O f(y) dy
 \end{aligned} \tag{44}$$

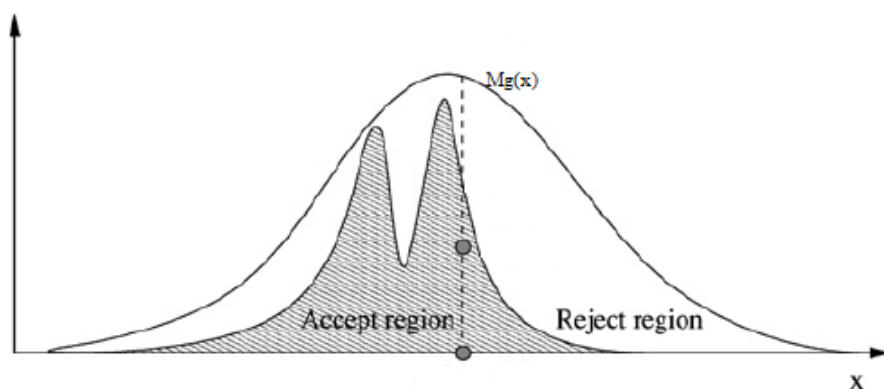
Nechť $X \sim f(x)$ a pro hustotu $g(x)$ platí $f(x) < Mg(x)$ pro nějaké $M \geq 1$ provedeme simulaci následovně:

Generujeme

$$Y \sim g(x), U|Y = y \sim \mathcal{U}_{(0, Mg(y))}$$

dokud $0 < u < f(y)$; položíme $x = y$.

Na první pohled je možné vidět některé nedostatky této metody, například variabilní nedeterministický čas do vygenerování N vzorků nebo požadavek $f(x) \leq m$. Díky tomuto požadavku potřebujeme určit konstantu m - de fakto „maximum“ distribuce $f(x)$. Dále je při vysoké „špičatosti“ $f(x)$ nebo špatné volbě m možný drastický nepoměr mezi počty přijetí a zamítnutí kandidátů na $X \sim f(x)$.



Obrázek 2: Ilustrace generalizované zamítací metody

6.1.3 MCMC, Metropolis-Hastings algoritmus

Markov Chain Monte Carlo vzorkovače jsou irreducibilní aperiodické markovské řetězce, které splňují podmínku detailní rovnováhy a mají stacionární distribuci, ze které chceme generovat vzorky.

Metropolis-Hastings (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller a Teller, 1953) algoritmus je jednou z více populárních MCMC metod. Prakticky používané MCMC algoritmy jsou často speciálními případy nebo rozšířeními tohoto algoritmu.

Krok M-H algoritmu spočívá ve vzorkování návrhové distribuce pro určení kandidáta x^* na následující stav, tento kandidát je na základě pravděpodobnosti přijetí A pak přijat nebo zamítnut, pokud je přijat markovský řetězec se posune do nového stavu, pokud je zamítnut řetězec zůstane ve stavu x .

Algoritmus:

1. inicializace náhodným nebo vhodně zvoleným stavem $x^{(0)}$
2. generujeme $u \sim \mathcal{U}_{(0,1)}$
3. generujeme $x^* \sim q(x^*|x^{(i)})$ - návrh následujícího stavu x^*
4. vyhodnocujeme pravděpodobnost přijetí nového stavu

$$A(x^{(i)}, x^*) = \min \left(1, \frac{p(x^*)q(x^{(i)}|x^*)}{p(x^{(i)})q(x^*|x^{(i)})} \right)$$

5. pokud $u < A(x^{(i)}, x^*)$ přijímáme nový návrh a $x^{i+1} = x^*$ jinak $x^{i+1} = x^{(i)}$
6. přejdeme na 2. a začneme novou iteraci, pokud ještě nemáme řetězec zvolené délky

Metropolis-Hastingsův algoritmus je velmi jednoduchý, ale je potřeba mít dobře zvolenou návrhovou distribuci $q(x^*|x^{(i)})$, která je zodpovědná za procházení vzorkované distribuce $p(x)$.

6.1.4 Simulované žíhání - numerická globální optimalizace

Předpokládejme nyní, že místo vzorkování distribuce chceme najít její globální maximum, jako například při hledání maximálního aposteriorního (MAP) odhadu nebo při konstrukci maximálně věrohodného (ML) odhadu.

Mohli bychom generovat markovský řetězec distribuce $p(x)$ a pak odhadnout globální extrém následovně:

$$\hat{x} \in \underset{x^{(i)}; i=1, \dots, N}{\text{Arg max}} p(x^{(i)}).$$

Nicméně toto není efektivním přístupem, protože vzorky $p(x)$ budou většinou relativně daleko od hledaného maxima, pokud rozdělení pravděpodobnosti není vysoce koncentrované kolem modu/maxima.

Kvalitní strategií pro hledání maxima je metoda simulovaného žíhání. Tato metoda je inspirována fyzikálním jevem při žíhání oceli (druh tepelného zpracování zahřátím a následným postupným pozvolným zchlazením).

Tato technika spočívá v simulaci nehomogenního markovského řetězce se stacionární distribucí $p^{1/T_i}(x)$, kde T_i je klesající hodnota „teploty“. Při velké teplotě jsou možné v markovském řetězci velké skoky, a takto se vyhýbáme uváznutí v lokálním maximu.

$$\hat{x} \in \underset{x^{(i)}; i=1, \dots, N}{\text{Arg max}} p^{1/T_i}(x^{(i)}).$$

Algoritmus:

1. inicializace náhodným nebo vhodně zvoleným stavem $x^{(0)}$
2. generujeme $u \sim \mathcal{U}_{(0,1)}$
3. generujeme $x^* \sim q(x^*|x^{(i)})$ - návrh následujícího stavu x^*
4. vyhodnocujeme pravděpodobnost přijetí nového stavu

$$A(x^{(i)}, x^*) = \min \left(1, \frac{p^{\frac{1}{T_i}}(x^*)q(x^{(i)}|x^*)}{p^{\frac{1}{T_i}}(x^{(i)})q(x^*|x^{(i)})} \right)$$

5. pokud $u < A(x^{(i)}, x^*)$ přijímáme nový návrh a $x^{i+1} = x^*$ jinak $x^{i+1} = x^{(i)}$
6. upravíme teplotu T_i podle zvolené strategie chlazení například: $T_i = (C \ln(i + T_0))^{-1}$, kde konstanty C a T_0 jsou vhodně zvolené pro daný problém
7. přejdeme na 2. a začneme novou iteraci, pokud ještě nemáme řetězec zvolené délky

7 PC implementace a řešení úloh

7.1 Řešená úloha

Při zátěžových zkouškách životnosti polovodičových součástek byla při různých teplotách získána následující data:

	Počet testovacích hodin	Počet selhání
$T_1 = 70^\circ C$	$t_1 = 7,247 \cdot 10^6$	$r_1 = 1$
$T_2 = 125^\circ C$	$t_2 = 10^7$	$r_2 = 3$

Tabulka 1: Data ke zkouškám životnosti

Vidíme, že se jedná o poissonův model, tedy r selhání v čase t . Konjugovaná apriorní hustota je pro tento případ dána funkcí gamma:

$$\pi(\lambda|\alpha, \beta) = \beta \cdot \Gamma(\beta\lambda; \alpha) \quad (45)$$

A podle bayesovy věty máme tedy aposteriorní rozdělení dáno jako:

$$\pi(\lambda|r; t, \alpha, \beta) = (t + \beta) \cdot \Gamma((t + \beta)\lambda; r + \alpha) \quad (46)$$

při pozorování r selhání v čase t . Pro konstrukci odhadu $\lambda \equiv \theta$ potřebujeme aposteriorní očekávanou hodnotu počtu selhání a aposteriorní rozptyl.

$$\begin{aligned} \hat{\lambda}_k &= \frac{(s + \alpha)}{(t + \beta)} \\ \text{Var} &= \frac{(s + \alpha)}{(t + \beta)^2} \end{aligned} \quad (47)$$

Kvadratická ztrátová funkce nám dává $\hat{\lambda}$ jako bayesovský estimátor, který srovnáme s odhadem získaným pomocí bezpečnostní ztrátové funkce s koeficientem $k = 1$.

$$\begin{aligned} \hat{\lambda}_b &= \frac{\sqrt{(s + \alpha) \cdot (s + \alpha + 1)}}{(t + \beta)} \\ &= \hat{\lambda}_k \cdot f \\ f &\equiv \sqrt{1 + \frac{1}{s + \alpha}} \end{aligned} \quad (48)$$

Parametry apriorního rozdělení si zvolíme na základě subjektivní pravděpodobnosti:

$$\begin{aligned} \alpha &= 103 \\ \beta &= 10^5 \end{aligned} \quad (49)$$

Vypočítáme koeficienty aposteriorního rozdělení pro zkoušky za teploty T_1 a T_2 :

$$\begin{aligned} r_1 + \alpha &= 104 & r_2 + \alpha &= 106 \\ t_1 + \beta &= 7,347 \cdot 10^6 & t_2 + \beta &= 1,01 \cdot 10^7 \end{aligned} \quad (50)$$

Na základě znalosti koeficientů aposteriorního Γ rozdělení:

$$\begin{aligned}\hat{\lambda}_{q1} &= 7,640 \cdot 10^8 & \hat{\lambda}_{q2} &= 1,07 \cdot 10^9 \\ \text{Var}_1 &= 5,613 \cdot 10^{15} & \text{Var}_2 &= 1,081 \cdot 10^{16}\end{aligned}\tag{51}$$

Bayesovský estimátor při bezpečnostní ztrátové funkci pak je:

$$\begin{aligned}\hat{\lambda}_{b1} &= \frac{\sqrt{2730}}{52} \cdot \hat{\lambda}_{q1} & \hat{\lambda}_{b2} &= \sqrt{\frac{107}{106}} \cdot \hat{\lambda}_{q2} \\ \hat{\lambda}_{b1} &= 7,676 \cdot 10^8 & \hat{\lambda}_{b2} &= 1,075 \cdot 10^9\end{aligned}\tag{52}$$

Je vidět, že odhad času do poruchy získaný při použití bezpečnostní ztrátové funkce, je v tomto případě, o zhruba 0,4 % vyšší oproti odhadu při použití kvadratické ztrátové funkce $\hat{\lambda}_q$.

7.2 Implementace algoritmů a numerická řešení

Výše uvedený příklad byl snadno analyticky řešitelný, nicméně v reálných problémech zpravidla činí vysoká dimenze problému analytické řešení prakticky neproveditelným. V takovýchto situacích vyvstává nutnost sáhnout po vhodných aproximacích a numerických metodách jako jsou metody typu Markov Chain Monte Carlo nebo Gibbsův vzorkovač. Tato problematika je detailněji diskutována v knihách a článcích [9, 12, 14, 15], ze kterých jsem také čerpal.

V této části tedy předvedeme implementaci Metropolis-Hastingsova algoritmu pro vzorkování rozdělení pravděpodobnosti a pro hledání globálního extrému. Obě varianty jsou poměrně jednoduché na implementaci. Pracujeme v prostředí Python 2.7.6. Pro grafický výstup byl zvolen balíček matplotlib. Pokud není řečeno jinak, ilustrační obrázky v této kapitole byly vytvořeny spuštěním uvedených kódů.

```

1  from math import *
2  from matplotlib.pylab import *
3  from random import gauss
4
5  def sdnorm(z):
6      #return exp(-(z*z/2.)/sqrt(2*pi))*N(0,1)
7      ny = [2,-2]
8      sigma = [3,3]
9      #return exp(-(z - ny[0])**2) /2.*sigma[0]**2/sqrt(2*pi)*sigma[0] #N(2,9)
10     return (0.5*(exp(-(z - ny[0])**2) /2.*sigma[0]**2/sqrt(2*pi)*sigma[0]) + 0.5*(exp(-(z - ny
        [1])**2) /2.*sigma[1]**2/sqrt(2*pi)*sigma[1])) #0.5 N(2,9) + 0.5N(-2,9)
11 n = 5000
12 alpha = 1
13 x = 0.
14 vec = []
15 vec.append(x)
16
17 #prop = uniform(-alpha,alpha,n) #random candidate for new x, uniform proposal distribution
18 for i in xrange(1,n):
19     can = x + gauss(0,1)#prop[i] #candidate
20     aprob = min([1.,sdnorm(can)/sdnorm(x)]) #acceptance probability, symmetric proposal density
21     u = uniform(0,1)
22     if u < aprob:
23         x = can
24         vec.append(x)
25     else:
26         vec.append(x)
27
28 #plotting the results :
29 # Markov Chain
30 x = arange(-5,5,.1)
31 y = sdnorm(x)
32 subplot(211)
33 title ('Metropolis-Hastings')
34 plot(vec)
35 subplot(212)
36
37 # Histogram
38 hist(vec, bins=50,normed=1)
39 plot(x,y, 'r--')
40 ylabel('Frequency')
41 xlabel('x')
42 legend(('PDF','Samples'))
43 show()

```

Výpis 1: Ukázka implementace M-H algoritmu v Pythonu

Poznamenejme, že ve výše uvedeném algoritmu využíváme faktu, že jsme si zvolili symetrickou návrhovou distribuci jako například uniformní distribuci nebo normální distribuci. Volba vhodné návrhové distribuce je velmi důležitá, protože přímo ovlivňuje kvalitu odhadů. Při nešikovné volbě návrhové distribuce (například s příliš úzkým rozptylem) by se mohlo stát, že náš řetězec uvízne v lokálním extrému, nebo že bude konvergovat

ke vzorkované distribuci příliš pomalu. Vhodnost návrhové distribuce je velice specifická charakteru vzorkovaného rozdělení pravděpodobnosti.

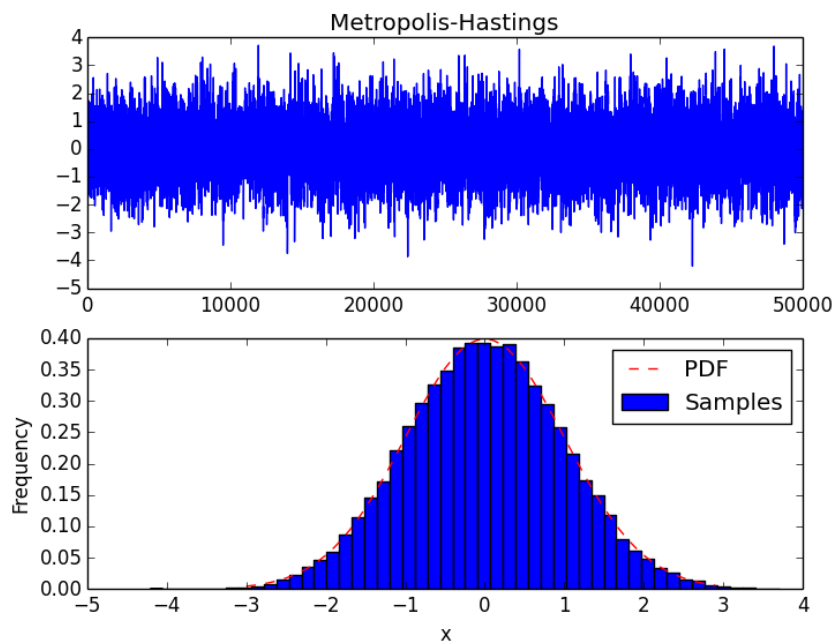
Díky symetrii návrhové distribuce:

$$q(x^{(i)}|x^*) = q(x^*|x^{(i)}) \quad (53)$$

Můžeme zjednodušit:

$$\frac{p(x^*)q(x^{(i)}|x^*)}{p(x^{(i)})q(x^*|x^{(i)})} = \frac{p(x^*)}{p(x^{(i)})} \quad (54)$$

V této situaci se jedná o speciální případ Metropolis-Hastingsova (1970) algoritmu, konkrétně Metropolisův algoritmus (1953).



Obrázek 3: Ukázka výstupu M-H algoritmu

```

1  from math import *
2  from matplotlib.pyplot import *
3  from random import gauss
4
5  T=50
6  C=4
7  def update_temperature(Tp,k):
8      T1=1/(C*log(k+T)); #
9      #T1=Tp/k
10     #print T1
11     return T1;
12
13  def sdnorm(z):
14     #return exp(-z*z/2.)/sqrt(2*pi)#N(0,1)
15     ny = [2,-2]
16     sigma = [3,3]
17     #return exp(-(z - ny[0])**2) /2.*sigma[0]**2/sqrt(2*pi)*sigma[0] #N(2,9)
18     return (0.7*(exp(-(z - ny[0])**2) /2.*sigma[0]**2/sqrt(2*pi)*sigma[0]) + 0.3*(exp(-(z - ny
19         [1])**2) /2.*sigma[1]**2/sqrt(2*pi)*sigma[1])) #0.5 N(2,9) + 0.5N(-2,9)
20
21  n = 200
22  Temp = T
23  alpha = 1
24  x = -2.
25  vec = []

```

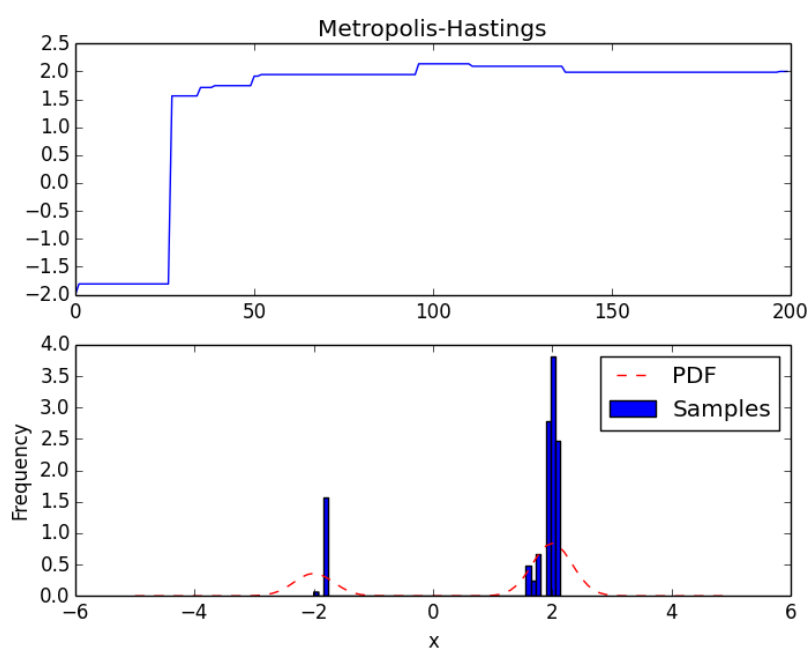
```

25  vec.append(x)
26
27  #prop = uniform(-alpha,alpha,n) #random candidate for new x, uniform proposal distribution
28  for i in xrange(1,n):
29      can = x + gauss(0,4)#prop[i] #candidate
30      aprob = min([1.,(sdnorm(can)**(1/Temp)/sdnorm(x)**(1/Temp))]) #acceptance probability,
          symmetric proposal density
31      u = uniform(0,1)
32      if u < aprob:
33          x = can
34          vec.append(x)
35      else:
36          vec.append(x)
37      Temp = update_temperature(Temp,i)
38
39  #plotting the results :
40  # Markov Chain
41  x = arange(-5,5,.1)
42  y = sdnorm(x)
43  subplot(211)
44  title ('Metropolis–Hastings')
45  plot(vec)
46  subplot(212)
47
48  # Histogram
49  hist(vec, bins=50,normed=1)
50  plot(x,y, 'r--')
51  ylabel('Frequency')
52  xlabel('x')
53  legend(('PDF','Samples'))
54  show()

```

Výpis 2: Ukázka implementace M-H algoritmu s úpravou pro simulování žíhání v Pythonu

Znovu poznamenejme, že ve výše uvedeném algoritmu využíváme faktu, že jsme si zvolili symetrickou návrhovou distribuci. V tomto algoritmu hraje důležitou roli chladicí strategie implementovaná funkcí `update_temperature`, která zodpovídá za konvergenci algoritmu a na začátku umožňuje flexibilnější procházení distribucí.



Obrázek 4: Ukázka výstupu M-H algoritmu s úpravou pro simulování žíhání

8 Závěr

Na závěr zrekapitulujeme témata jež byla v práci zpracována.

Uvedli a dokázali jsme Bayesovu větu. V této práci byla prezentována teorie rozhodování za nejistoty, možnosti volby ztrátové funkce a jejich přidružené bayesovské estimátory.

Byly představeny metody Markov Chain Monte Carlo, které mají široké pole uplatnění.

Při rozhodování pod downside damaging ztrátovou funkcí dostáváme vyšší odhady než při kvadratické ztrátové funkci, což je při bližším pohledu na tvar obou ztrátových funkcí jasné. Nicméně jsem očekával, že u příkladu zkoušek životnosti polovodičových součástek nedojde k výraznější změně odhadu.

Zahrnutí teorie týkající se cenzorovaných dat a markovských řetězců by bylo přirozeným rozšířením této práce, stejně jako prohloubení kapitoly věnované numerickému výpočtu integrálů zahrnutím teoretického základu Gibbsova vzorkovače a následně i jeho algoritmu. MCMC algoritmy jsou také relativně snadno paralelizovatelné takže kapitola věnovaná jednoduché implementaci paralelizovaného algoritmu by jistě práci velice prospěla.

9 Reference

- [1] BERGER, J. O., *Statistical Decision Theory and Bayesian Analysis*, 1988, Springer.
- [2] MARTZ H. F., WALLER R. A., *Bayesian Reliability Analysis*, 1982, Wiley, ISBN 0-471-86425-0 .
- [3] NOSTRØM, J. G., *Bayesian analysis & decision theory in reliability and risk analysis*, únor 1994, Diplomová práce na The Norwegian University of Science and Thechnology.
- [4] NOSTRØM, J. G., *The Use of Precautionary Loss Function in Risk Analysis*, IEEE Transactions on Reliability, vol. 45, No. 3, 1996.
- [5] BRIŠ R., LITSCHMANNOVÁ M., *STATISTIKA II*, E-learningový prvek pro podporu výuky odborných a technických předmětů, v rámci projektu CZ.O4.01.3/3.2.15.2/0326, VŠB TU Ostrava, 2007, ISBN 978-80-248-1482-7.
- [6] BRIŠ R., Bayes approach in RDT using accelerated and long-term life data, V *Reliability Engineering & System Safety*, Volume 67, Issue 1, leden 2000, strana 9–16
- [7] SMITH J. Q., *Bayesian Decision Analysis: Principles and Practice*, Cambridge University Press (23 Sep 2010), ISBN-10: 0521258723, ISBN-13: 978-0521258722
- [8] JAYANTA K. GHOSH, DELAMPADY Mohan, *An introduction to Bayesian analysis theory and methods.*, New York, NY: Springer Verlag New York, 2006. ISBN 03-873-5433-6.
- [9] DEVROYE, Luc. *Non-uniform random variate generation*. New York: Springer-Verlag, 1986. ISBN 03-879-6305-7.
- [10] HUŠKOVÁ Marie, *Bayesovské metody*, Skripta, Univerzita Karlova v Praze
- [11] DOWNEY Allen B., *Think Bayes*, Green Tea Press, 2012, ISBN-10: 1449370780, ISBN-13: 978-1449370787
- [12] ANDRIEU Christophe , DE FREITAS Nando , DOUCET Arnaud , JORDAN Michael I., An Introduction to MCMC for Machine Learning, In *Machine Learning* Leden 2003, Volume 50, Issue 1-2, pp 5-43.
- [13] BERNARDO, José M., *Bayesian theory*. Chichester: John Wiley, c2000, xiv, 586 s. ISBN 04-714-9464-X.
- [14] LIANG, F, Chuanhai LIU a Raymond J CARROLL. *Advanced Markov chain Monte Carlo methods: learning from past samples*. Chichester, West Sussex, U.K.: Wiley, 2010, xix, 357 p. Wiley series in computational statistics. ISBN 04-707-4826-5.
- [15] KENDALL, W, F LIANG a J WANG. *Markov chain Monte Carlo: innovations and applications*. Hackensack, NJ: World Scientific, c2005, xviii, 220 p. Lecture notes series (National University of Singapore. Institute for Mathematical Sciences), v. 7. ISBN 978-981-2564-276.

- [16] KRACÍK Jan *Bayesovské metody*, doprovodný text k předmětu Statistika III
- [17] SAVAGE, L. J. *The subjective basis of statistical practice*, Technical report, Department of Statistics, University of Michigan, Ann Arbor.